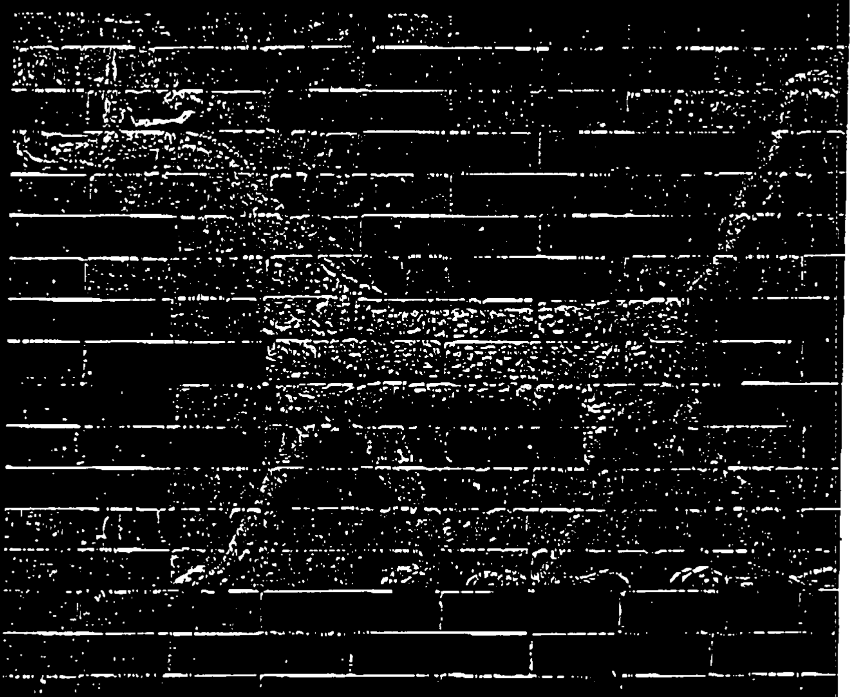
WILEY-VCH

Roberto Todeschini, Viviana Consonni

Handbook of Molecular Descriptors



Methods and Principles in Medicinal Chemistry

Volume 11

Edited by
R. Mannhold,
H. Kubinyi,
H. Timmerman

ms, derived
le distances
g the diago-
obtained by
fields in grid
ely positive-
; resulting in
is that their
993b; van de

93] [Zakarya
996] [Patter-
d Legendre,

autocorrela-

on density)

on density)

Balaban dis-

ptors (⊙ elec-

on)

ectron density)

iptors

es

B

backward Fukui function → **quantum-chemical descriptors** (⊙ Fukui function)

Balaban centric indices → **centric indices**

Balaban distance connectivity index → **Balaban distance connectivity indices**

Balaban distance connectivity indices

The formerly proposed and the most important of this series of topological indices is the **Balaban distance connectivity index** J (also called **distance connectivity index** or **average distance sum connectivity**). It is one of the most discriminating → *molecular descriptors* and its values do not increase substantially with molecule size or number of rings; it is defined in terms of sums over each ith row of the → *distance matrix* **D**, i.e. the → *vertex distance degree* σ [Balaban, 1982; Balaban, 1983a]. It is defined as:

$$J = \frac{B}{C+1} \cdot \sum_b \left(\sigma_i \cdot \sigma_j\right)_b^{-1/2} = \frac{1}{C+1} \cdot \sum_b \left(\bar{\sigma}_i \cdot \bar{\sigma}_j\right)_b^{-1/2}$$

where σ_i and σ_j are the vertex distance degrees of two adjacent atoms, and the sum runs over all the molecular bonds b; B is the number of bonds in the molecular graph G, and C, called the → *cyclomatic number*, the number of rings. The denominator $C + 1$ is a normalization factor against the numbes of rings in the molecule. $\bar{\sigma}_i = \sigma_i / B$ is the **average vertex distance degree**; it was observed that within an isomeric series the average distance degrees are low in the more branched isomers.

To further improve the discriminant power of the Balaban index J, a set of new LOVIs was defined as:

$$t_i = \frac{\sigma_i}{\delta_i}$$

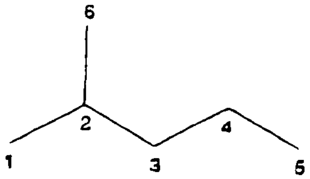
where δ_i is the ith → *vertex degree*. Therefore, the **J$_t$ index** was defined as:

$$J_t = \frac{B}{C+1} \cdot \sum_b \left(t_i \cdot t_j\right)_b^{-1/2}$$

The idea behind these LOVIs is that usually the vertices with the highest distance sums have the lowest vertex degrees, thus enhancing the intramolecular differences [Balaban, 1994a].

The J index for multigraphs is calculated by the distance sums of the → *multigraph distance matrix* where the distances are obtained by weighting each edge with the inverse of its → *conventional bond order* (→ *relative topological distance*); the sum runs over all pairs of adjacent vertices and B is the number of edges in the graph without accounting for their multiplicity.

Example : 2-methylpentane



Distance matrix **D**

Atom	1	2	3	4	5	6	σ_i
1	0	1	2	3	4	2	12
2	1	0	1	2	3	1	8
3	2	1	0	1	2	2	8
4	3	2	1	0	1	3	10
5	4	3	2	1	0	4	14
6	2	1	2	3	4	0	12

Vertex degrees

Atom	δ_i	t_i
1	1	12
2	3	2.667
3	2	4
4	2	5
5	1	14
6	1	12

$$J = \frac{B}{C+1} \cdot \left[(\sigma_1 \cdot \sigma_2)^{-1/2} + (\sigma_6 \cdot \sigma_2)^{-1/2} + (\sigma_2 \cdot \sigma_3)^{-1/2} + (\sigma_3 \cdot \sigma_4)^{-1/2} + (\sigma_4 \cdot \sigma_5)^{-1/2} \right] =$$
$$= 5 \cdot \left[(12 \cdot 8)^{-1/2} + (12 \cdot 8)^{-1/2} + (8 \cdot 8)^{-1/2} + (8 \cdot 10)^{-1/2} + (10 \cdot 14)^{-1/2} \right] = 2.6272$$

$$J_t = \frac{B}{C+1} \cdot \left[(t_1 \cdot t_2)^{-1/2} + (t_6 \cdot t_2)^{-1/2} + (t_2 \cdot t_3)^{-1/2} + (t_3 \cdot t_4)^{-1/2} + (t_4 \cdot t_5)^{-1/2} \right] =$$
$$= 5 \cdot \left[(12 \cdot 2.667)^{-1/2} + (12 \cdot 2.667)^{-1/2} + (2.667 \cdot 4)^{-1/2} + (4 \cdot 5)^{-1/2} + (5 \cdot 14)^{-1/2} \right] = 5.0141$$

Box B-1.

In order to account for both bond multiplicity and heteroatoms, **Balaban modified distance connectivity indices** J^X and J^Y were proposed [Balaban, 1986a; Balaban *et al.*, 1990a]. These are defined in the same way as the Balaban distance connectivity index but derived from the → *multigraph distance matrix* ***D** instead of the original distance matrix **D**:

$$J^X = \frac{B}{C+1} \cdot \sum_b \left({}^*\sigma_i^X \cdot {}^*\sigma_j^X \right)^{-1/2}$$
$$J^Y = \frac{B}{C+1} \cdot \sum_b \left({}^*\sigma_i^Y \cdot {}^*\sigma_j^Y \right)^{-1/2}$$

where B is the bond number, C is the cyclomatic number, and the sum runs over all bonds b in the graph, each being weighted by the inverse square root of the product of the → *multigraph distance degree* of the incident vertices. The distance degrees are calculated as:

=

0141

dified
et al.,
index
tance

er all
uct of
s are

$$^{*}\sigma_i^X = X_i \cdot {}^{*}\sigma_i = X_i \cdot \sum_{j=1}^{A} [{}^{*}\mathbf{D}]_{ij} \quad \text{and} \quad X_i = 0.4196 - 0.0078 \cdot Z_i + 0.1567 \cdot L_i$$

$$^{*}\sigma_i^Y = Y_i \cdot {}^{*}\sigma_i = Y_i \cdot \sum_{j=1}^{A} [{}^{*}\mathbf{D}]_{ij} \quad \text{and} \quad Y_i = 1.1191 + 0.0160 \cdot Z_i - 0.0537 \cdot L_i$$

where $^{*}\sigma$ is the vertex distance degree calculated on multigraph distance matrix $^{*}\mathbf{D}$, the quantities X and Y are recalculated atomic Sanderson electronegativities and covalent radii relative to carbon atom, obtained as a function of the atomic number Z_i and the principal quantum number L_i of the atom; for atoms different from B, C, N, O, F, Si, P, S, Cl, As, Se, Br, Te, and I the X and Y values are set at one. X and Y indices account for the presence of heteroatoms in the molecule.

Another generalization of the Balaban index J, so as to account for heteroatoms in the molecule, is the → *Barysz index* calculated on the → *Barysz distance matrix*.

→ *JJ indices* derived from the → *Wiener matrix* were proposed as a generalization of the Balaban index in analogy with the Kier-Hall → *connectivity indices*.

The **3D-Balaban index** ^{3D}J was derived from the → *geometry matrix* G as:

$$^{3D}J = \mathcal{IB}(\mathbf{G}) = \frac{B}{C+1} \cdot \sum\nolimits_b \left({}^{G}\sigma_i \cdot {}^{G}\sigma_j\right)_b^{-1/2}$$

where $\mathcal{IB}$ is the → *Ivanciuc-Balaban operator*, $^{G}\sigma_i$ and $^{G}\sigma_j$ are the → *geometric distance degree* of the two vertices incident with the b bond [Mihalic *et al.*, 1992a].

A **Balaban-type index** *DJ* [Balaban and Diudea, 1993] was defined as:

$$DJ = \sum_{i=1}^{A} dj_i = \sum_{i=1}^{A} \sum\nolimits_{j \in V_{i1}} \left(\frac{\sigma_i}{w_i(1+f_i)} \cdot \frac{\sigma_j}{w_j(1+f_j)} \right)^{-1/2}$$

where A is the → *atom number*, f is the → *multigraph factor*, w is a weighting factor accounting for heteroatoms, and the inner sum runs over all vertices j at distance 1 from the ith atom, i.e. vertices bonded to the ith atom; dj are local vertex invariants accounting for heteroatoms and bond multiplicity. When the factor w is equal to one and the multigraph factor is equal to zero then the index DJ is related to the Balaban index J by the following:

$$DJ = 2 \cdot J \cdot \frac{C+1}{B}$$

📖 [Balaban and Quintar, 1983] [Barysz *et al.*, 1983a] [Balaban and Filip, 1984] [Balaban *et al.*, 1985e] [Sabljic, 1985] [Mekenyan *et al.*, 1987] [Balaban and Ivanciuc, 1989] [Balaban *et al.*, 1990b] [Balaban *et al.*, 1992a] [Nikolic *et al.*, 1993a] [Guo and Randic, 1999]

Balaban ID number → **ID numbers**

Balaban modified distance connectivity indices → **Balaban distance connectivity indices**

Balaban-type index → **Balaban distance connectivity indices**

Bartell resonance energy → **resonance indices**

barycentre : *centre of mass* → **centre of a molecule**

Barysz index → **weighted matrices**

Barysz distance matrix → **weighted matrices**

Kekulé structure count : *Kekulé number*

Kellog and Abraham interaction field → molecular interaction fields
(⊙ hydrophobic fields)

Kier alpha-modified shape descriptors → Kier shape descriptors

Kier bond rigidity index → flexibility indices

Kier-Hall connectivity indices : *connectivity indices of mth order* → **connectivity indices**

Kier-Hall connectivity matrix → weighted matrices

Kier-Hall electronegativity → vertex degree

Kier-Hall solvent polarity index → electric polarization descriptors

Kier molecular flexibility index → flexibility indices

Kier steric descriptor → steric descriptors

Kier shape descriptors (κ)

Topological shape descriptors ${}^m\kappa$ defined in terms of the number of graph vertices A and the number of paths ${}^mP$ with length m ($m = 1,2,3$) in the → *H-depleted molecular graph*, according to the following:

$$^1\kappa = 2 \cdot \frac{{}^1P_{\max} \cdot {}^1P_{\min}}{({}^1P)^2} = \frac{A(A-1)^2}{({}^1P)^2} \qquad {}^2\kappa = 2 \cdot \frac{{}^2P_{\max} \cdot {}^2P_{\min}}{({}^2P)^2} = \frac{(A-1)(A-2)^2}{({}^2P)^2}$$

$$^3\kappa = 4 \cdot \frac{{}^3P_{\max} \cdot {}^3P_{\min}}{({}^3P)^2} = \begin{cases} \dfrac{(A-3)(A-2)^2}{({}^3P)^2} & \text{for even } A \ (A>3) \\ \dfrac{(A-1)(A-3)^2}{({}^3P)^2} & \text{for odd } A \ (A>3) \end{cases}$$

where ${}^mP_{\min}$ and ${}^mP_{\max}$ are the minimum and maximum mth order → *path count* in the molecular graphs of molecules with the same → *atom number A* [Kier, 1985; Kier, 1986b]. These extremes are obtained from two reference structures chosen in an isomeric series and, for the ith molecule, is therefore:

$$^mP_{\min} \leq {}^mP_i \leq {}^mP_{\max}$$

The reference structure for ${}^1P_{\min}$ is the → *linear graph* while for ${}^1P_{\max}$ it is the → *complete graph* in which all atoms are bonded to each other; their numerical values are calculated as follows:

$$^1P_{\min} = A - 1 \qquad {}^1P_{\max} = \frac{A(A-1)}{2}$$

The scaling factor of 2 in the numerator of ${}^1\kappa$ index formula makes the value ${}^1\kappa = A$ when there are no cycles in the graph of the molecule. Monocyclic molecules have a lower value and bicyclic structures have an even lower value. The structural information encoded in ${}^1\kappa$ is related to the complexity, or more precisely, the number of cycles of a molecule.

The reference structure for ${}^2P_{\min}$ is the linear graph, while for ${}^2P_{\max}$ it is the → *star graph*, in which all atoms but one are adjacent to a central atom; their numerical values are calculated as follows;

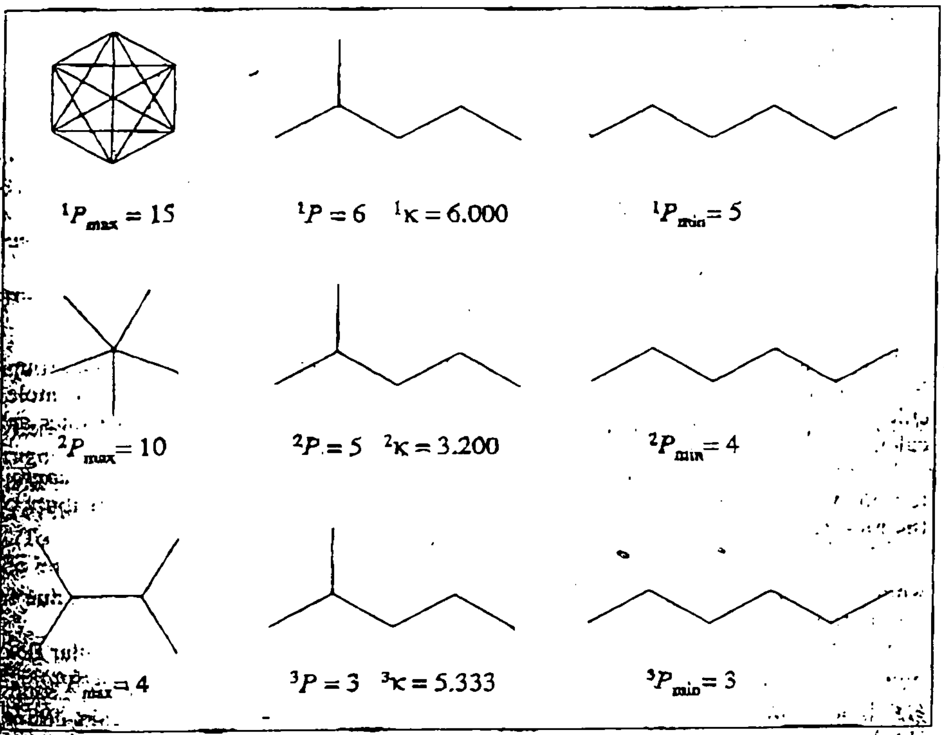
$$^2P_{\min} = A - 2 \qquad {}^2P_{\max} = \frac{(A-1)(A-2)}{2}$$

where A is the total number of vertices in the graph. The scaling factor of 2 in the numerator of $^2\kappa$ index formula makes the value $^2\kappa = A - 1$ for all linear graphs. The information encoded by $^2\kappa$ index is related to the degree of star graph-likeness and linear graph-likeness, i. e. $^2\kappa$ encodes information about the spatial density of atoms in a molecule.

The reference structure for $^3P_{\min}$ is the linear graph while for $^3P_{\max}$ it is the *twin star graph*; their numerical values are calculated as follows:

$$^3P_{\min} = A - 3 \qquad ^3P_{\max} = \begin{cases} \dfrac{(A-2)^2}{4} & \text{for even } A \\ \dfrac{(A-1)(A-3)}{4} & \text{for odd } A \end{cases}$$

The scaling factor of 4 is used in the numerator of $^3\kappa$ index to bring $^3\kappa$ onto approximately the same numerical scale as the other kappa indices. The $^3\kappa$ values are larger when → *molecular branching* is nonexistent or when it is located at the extremities of a graph; $^3\kappa$ encodes information about the centrality of branching.



Box K-1.

To take into account the different shape contribution of heteroatoms and hybridization states, **Kier alpha-modified shape descriptors** $^m\kappa_\alpha$ (m = 1,2,3) were proposed [Kier, 1986a] by the following:

$$^{1}\kappa_{\alpha} = \frac{(A+\alpha)(A+\alpha-1)^2}{(^{1}P+\alpha)^2} \qquad ^{2}\kappa_{\alpha} = \frac{(A+\alpha-1)(A+\alpha-2)^2}{(^{2}P+\alpha)^2}$$

$$^{3}\kappa_{\alpha} = \begin{cases} \dfrac{(A+\alpha-3)(A+\alpha-2)^2}{(^{3}P+\alpha)^2} & \text{for even } A \ (A>3) \\ \dfrac{(A+\alpha-1)(A+\alpha-3)^2}{(^{3}P+\alpha)^2} & \text{for odd } A \ (A>3) \end{cases}$$

where α is a parameter derived from the ratio of the → *covalent radius* R_i of the *i*th atom relative to the sp³ carbon atom (R_{Csp3}):

$$\alpha = \sum_{i=1}^{A} \left(\frac{R_i}{R_{\mathrm{Csp^3}}} - 1 \right)$$

The only non-zero contributions to α are given by heteroatoms or carbon atoms with a valence state different from sp³ (Table K-1).

Table K-1. Covalent radius *R* and α parameter values.

Atom / Hybrid	*R* (Å)	α	*Atom / Hybrid*	*R* (Å)	α
C_{sp3}	0.77	0	P_{sp3}	1.10	0.43
C_{sp2}	0.67	−0.13	P_{sp2}	1.00	0.30
C_{sp}	0.60	−0.22	S_{sp3}	1.04	0.35
N_{sp3}	0.74	−0.04	S_{sp2}	0.94	0.22
N_{sp2}	0.62	−0.20	F	0.72	−0.07
N_{sp}	0.55	−0.29	Cl	0.99	0.29
O_{sp3}	0.74	−0.04	Br	1.14	0.48
O_{sp2}	0.62	−0.20	I	1.33	0.73

Kappa indices can also be calculated for molecular fragments and functional groups X. The calculation of these indices for groups was performed using a "pseudo-molecule" X–X: two fragments X of the same kind are linked together, kappa values are calculated for the pseudo-molecule and this is then divided by two.

In order to quantify the shape of the whole molecule, Kier proposed a linear combination of the above defined κ indices, each representing a particular shape attribute of the molecule:

$$shape = b_0 \cdot {}^{0}\kappa + b_1 \cdot {}^{1}\kappa + b_2 \cdot {}^{2}\kappa + b_3 \cdot {}^{3}\kappa$$

where ⁰κ is the → *Kier symmetry index* used to encode the shape contributions due to symmetry.

Specific combinations of κ indices were also proposed as indices of molecular flexibility (→ *Kier molecular flexibility index*) and steric effects (→ *Kier steric descriptor*).

📖 [Kier, 1986c] [Kier, 1987a] [Kier, 1987b] [Kier, 1987c] [Gombar and Jain, 1987a] [Mokrosz, 1989] [Kier, 1990] [Hall and Kier, 1991] [Skvortsova *et al.*, 1993] [Kier, 1997] [Hall and Vaughn, 1997b]

Kier symmetry index → **symmetry descriptors**

***K*_Z index** → **Hosoya Z matrix**

Appendix C. Software

Some packages explicitly related to the calculation of the molecular descriptors for QSAR/QSPR are collected below, in alphabetic order.

General programs for computational quantum-chemistry, molecular modelling and log*P* calculations are not explicitly considered in this list. An extended list of computational chemistry programs can be found at the WebSite http://www.netsci.org/Resources/Software/.

ADAPT	Prof. P.C. Jurs, PennState University, University Park, PA 16802, USA
Description:	A QSAR toolkit with descriptor generation (topological, geometrical, electronic, and physicochemical descriptors), variable selection, regression and artificial neural network modelling.
Reference:	[Jurs *et al.*, 1979]
WebSite:	http://zeus.chem.psu.edu/

ASP	Oxford Molecular Ltd., Oxford Science Park, Oxford OX4 4GA, UK
Description:	Calculates a quantitative measure of molecular similarity based on molecule alignment, shape and electronic properties. Within TSAR 3D package.
WebSite:	http://www.oxmol.com/

CERIUS^2	Molecular Simulations Inc. – 9685 Scranton Road, San Diego, CA 92121-7352, USA
Description:	C2-Descriptors+ provides a range of generic descriptors, describing topological, electronic, and structural features.
WebSite:	http://www.msi.com/life/products/cerius2/modules/descriptor.html

CODESSA	Semichem Inc. – 7204 Mullen, Shawnee, KS 66216, USA
Description:	Calculation of several topological, geometrical, constitutional, thermodynamic, electrostatic, and quantum-chemical descriptors, including tools for regression modelling and variable selection.
Reference:	[Katritzky *et al.*, 1995]
WebSite:	http://www.semichem.com/

DRAGON	Prof. R. Todeschini – distributed by Talete srl, via Pisani 13, 20124 Milano, Italy
Description:	Calculation of several sets of molecular descriptors from molecular geometries (topological, geometrical, WHIM, 3D-MoRSE, molecular profiles, etc.).
WebSite:	http://www.disat.unimib.it/chm/

GRIN/GRID	Molecular Discovery Ltd. – West Way House, Elms Parade, Oxford OX2 9LL, UK
Description:	Calculates the GRID empirical force field at grid points. Last release: V.11 – 1993
Reference:	[Goodford, 1985]

HQSAR	Tripos Inc. – 1699 South Hanley Rd., St.Louis, MO 63144-2913, USA
Description:	A part of the SYBYL environment providing hologram descriptors.
WebSite:	http://www.tripos.com/

HYBOT-PLUS	Prof. O. Raevsky – Russian Academy of Science, IPAC
Description:	Calculation of hydrogen bond and free energy factors.
Reference:	[Raevsky, 1997]
WebSite:	http://www.ipac.ac.ru/qsar/index.htm

HYPERCHEM 6	Hypercube, Inc. – 1115 NW 4th Street, Gainsville, FL 32601, USA
Description:	Calculation of optimized geometries with several computational methods, also providing total surface area, molecular volume, molar refractivity, log P, polarizability and atomic charges. Last release: 6
WebSite:	http://www.hyper.com/

MOLCONN-Z	Prof. L.H. Hall – 2 Davis Street, Quincy, MA 02170, USA
Description:	Successor of MOLCONN-X, MOLCONN-Z calculates the most well-known topological descriptors, including electrotopological and orthogonalized indices. Last release: 3.0
WebSite:	http:// www.eslc.vabiotech.com/molconn/manuals/310s/preface1.html

MULTICASE	Multicase Inc. – PO 22517, Beachwood, OH 44122, USA
Description:	Prediction of biological activities by substructure descriptors.
Reference:	[Klopman, 1992]
WebSite:	http://www.multicase.com/

OASIS	Prof. O. Mekenyan – Bourgas University, 8010 Bourgas, Bulgaria
Description:	Calculation of steric, electronic, and hydrophobic descriptors.
Reference:	[Mekenyan *et al.*, 1990a]
WebSite:	http://omega.btu.bg/~omekenya/

PETRA	Molecular Networks GmbH – Langemarckplatz 1, D-91054 Erlangen (Germany)
Description:	Empirical methods for the calculation of charges and bond energies for use in QSAR.
Reference:	[Gasteiger, 1988; Löw and Saller, 1988]

POLLY	Prof. S. Basak – Minnesota University of Duluth, 5013 Miller Trunk Highway, Duluth, MN 55811, USA
Description:	Calculation of topological connectivity indices. Last release: 2.3
Reference:	[Basak *et al.*, 1988a]

hods,
ty, log P,

ll-known
ized indices

d

ı (Gormany)
s for use in

ık Highway,

SciQSAR 2D	SciVision – 200 Wheeler Road, Burlington, MA 01803, USA
Description:	Calculation of several topological molecular descriptors (connectivity, shape, electrotopological descriptors).
WebSite:	http://www.scivision.com/

SYBYL/QSAR	Tripos Inc. – 1699 South Hanley Rd., St.Louis, MO 63144-2913, USA
Description:	SYBYL module for the calculation of EVA descriptors, CoMFA and CoMSIA fields, also including several QSAR tools. Last release: 6.1
WebSite:	http://www.tripos.com/

TSAR	Oxford Molecular Ltd., Oxford Science Park, Oxford OX4 4GA, UK
Description:	Statistical and database functions with molecular and substituent property calculations. Within TSAR 3D package.
Reference:	http://www.oxmol.com/